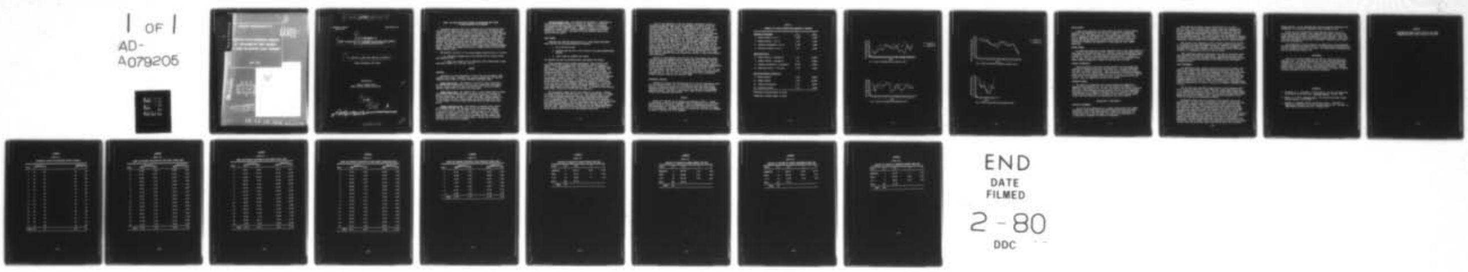
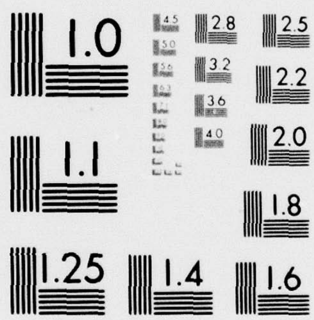
AD-A079 205 ADJUTANT GENERAL'S OFFICE (ARMY) WASHINGTON DC F/G 5/10
INTER- AND INTRA-APPARATUS VARIANCE OF PSYCHOMOTOR TEST SCORES --ETC(U)
MAY 61 D A DOBBINS , H MARTINEK
UNCLASSIFIED AGO-HFRB-RM-61-5 NL

1 OF 1
AD-
A079205



END
DATE
FILMED
2-80
DDC



MICROCOPY RESOLUTION TEST CHART

NATIONAL BUREAU OF STANDARDS-1963-A

(14) AGO-AFRB-RM-61-5

Army Project Number
2195-60-001

Helicopter b-41

(9)

Research Memor[redacted] 61-5

(6)

INTER- AND INTRA-APPARATUS VARIANCE OF PSYCHOMOTOR TEST SCORES OF ARMY HELICOPTER PILOT TRAINEES.

(10)

D. A. Dobbins, H. Martinek, [redacted] Alan A. Anderson

Nathan Rosenberg, Task Leader

Submitted by

Samuel H. King, Chief
Combat Systems Research Laboratory

Accession For	
NTIS GRA&I	✓
DDC TAB	
Unannounced	
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or special
A	

(11)

May [redacted]61

(12) 21

~~This document is intended for use only by The Adjutant General, AGTL, and is not available for distribution.~~

003650

INTER- AND INTRA-APPARATUS VARIANCE OF PSYCHOMOTOR TEST SCORES OF ARMY HELICOPTER PILOT TRAINEES

Psychomotor apparatus tests were introduced into the experimental selection battery for Army helicopter pilot trainees because of the demonstrated validity of such tests in the selection of Air Force pilots for fixed-wing aircraft. Increments in validity gained from psychomotor tests, however, must be balanced against several disadvantages inherent in their use. Among these disadvantages are expense, need for centralized testing facilities, need for continuing maintenance, possible lack of score comparability from machine to machine, and score fluctuations when the same machine is used over a period of time. The study reported here was concerned with score differences from machine to machine and with progressive score changes during continuing machine usage. The study was a partial replication of several World War II Air Force studies (Melton, 1947).

The specific objectives of the present Research Memorandum are to describe

1. Differences between mean scores derived from two copies of the same apparatus test. and

2. Mean score changes of four psychomotor tests administered to Army helicopter classes over a two year period.

METHOD

APPARATUS

Following is a brief description of each of the four psychomotor tests included in the study. Diagrams, more complete descriptions, and validity summaries may be found in a recent reference (Fleishman, 1956).

Rotary Pursuit Test. The examinee is required to keep a prod-stylus in contact with a small metallic disk set into a larger disk revolving at 60 rpm. Testing consists of three 20-second trials. The sum of contact time over three trials was used as the total score.

Rudder Control Test. The examinee sits in a mock cockpit which is thrown off balance by his own weight unless he applies the proper correction by means of foot pedals. His task is to keep the cockpit lined up with one of three target lights located on a panel before him. Testing consists of one 90-second trial with a center target, and one 348-second trial with a triple target. The sum of scores for both trials was used as the total score.

Complex Coordination Test. The examinee is presented with a double row of lights (one red row, one green row) in the approximate pattern of an "H" on its side. One red light in each row is lit. The examinee is required to match the position of a stimulus light in each of three dimensions by adjustments of the stick and rudder controls. After each matching, a new pattern of red lights is presented and the examinee must reproduce this pattern with the green lights. Testing time is 8 minutes. The total score was used for analysis.

Direction Control Test. The examinee must manipulate a combination of switches and buttons as rapidly as possible in response to a series of visual patterns differing from one another with respect to the spatial arrangement of their component parts. Emphasis is placed on selecting an appropriate response where a complex sequence of movements must be made. Testing time is 8 minutes. The number of correct patterns achieved was used as the total score.

GROUP TESTED

Examinees were enlisted applicants for U. S. Army Primary Helicopter School who met the following preselection standards:

1. 20 to 30 years of age.

2. Aptitude area GT score of 110 or above on the Army Classification Battery.

3. Class I medical standards for flying.

All examinees who met the preceding three requirements were tested.

Three psychomotor tests--Complex Coordination, Rotary Pursuit, and Rudder Control--were administered to 758 examinees from 16 successive classes. A fourth psychomotor test, Direction Control, was administered to 288 examinees, representative of the first seven consecutive classes. The first nine classes were tested at Fort Rucker, Alabama; the last seven classes were tested at Camp Wolters, Texas. Testing took place during 1956 and 1957. The number of examinees in consecutive classes varied from 22 to 72. Within any one class, however, an equal number of different examinees was assigned at random to Apparatus A and B. This requirement led to the random elimination of some cases. All examinees tested with Apparatus A for any test were the same examinees tested with Apparatus A for all other tests; the same is true for examinees tested with Apparatus B. In the case of the Rotary Pursuit, Rudder Control and Complex Coordination tests, the total number tested with each apparatus was 379. In the case of the Direction Control Test, 144 men were tested with each apparatus.

The experimental design used for this study was not optimal. Ideally, all subjects would have been pre-tested on a third apparatus copy or a highly correlated psychomotor paper-and-pencil test, thus equating Groups A and B as to initial psychomotor ability. Subsequent differences found between scores on Apparatus A and B could then be attributed with more certainty to true apparatus differences. However, the present study was initiated long after data collection had begun for the helicopter pilot selection study when a decrement--apparently progressive--in mean score on the Complex Coordination test was noted. Apparatus differences obtained were thus confounded with initial ability differences in the random-groups design used.

Since it was impossible to exercise adequate experimental controls, a subsequent attempt was made to test the hypothesis of initial ability differences among groups. The index of initial ability used was failure or success in later helicopter training. While this index is admittedly gross and contains components other than psychomotor ability, it was the only appropriate measure readily available. Accordingly, the proportions passing and failing helicopter training were subjected to a chi-square analysis. Pass-fail data are shown in Table A-1 of the Appendix. The proportion tested under Apparatus A, who later passed the helicopter training program, was .49--for those tested under Apparatus B, .50. The chi square, evaluated with one degree of freedom, was .08, which is not statistically significant ($P > .70$). A 2 x 16 chi square was then computed to test for class differences. The chi square, evaluated with 15 degrees of freedom, was 16.24, which was also not statistically significant ($P > 30$). Comparable chi square analyses of pass-fail data were made for the first seven classes on which Direction Control scores were available. Chi squares of .12 ($df = 1$; $P > .70$) and .46 ($df = 6$; $P > .99$) were found for apparatus and class respectively, neither of which were statistically significant.

Results of these analyses, plus the fact that large numbers were tested with each apparatus, suggested that the samples used in the present study were initially reasonably equated in ability to pass helicopter training. It was further assumed that the samples were at least grossly similar with respect to initial psychomotor differences by virtue of the known psychomotor representation in the helicopter training criteria. The Rotary Pursuit, Rudder Control, Direction Control, and Complex Coordination Tests correlated .29, .11, .17, and .28, respectively, with performance in the pre-flight phase of helicopter pilot training, and .30, .39, .15, and .41 respectively with flight training performance (Zeidner, Martinek, and Anderson, 1958).

STATISTICAL ANALYSIS

Analyses of variance were performed for each of the four psychomotor tests. For each psychomotor test, one 2 x 16 analysis was performed with two main effects, a first-order interaction and a residual error term within groups. Differences between the means of two copies of the same test were designated apparatus effects. Variations in mean scores among successive classes (time trends) were designated class effects.

RESULTS

Values of F obtained in all analyses are presented in Table 1. Trends in performance on each of the psychomotor tests (by apparatus) over the sixteen classes are shown in Figures 1 through 4. Means and standard deviations for each psychomotor test, by apparatus and class, are shown in Tables A-2 through A-5 of the Appendix. Individual analyses of variance for each test are presented in Tables A-6 through A-9 of the Appendix.

Table 1

SUMMARY OF F-VALUES DERIVED FROM ANALYSES OF VARIANCE

	F	df
Apparatus Differences		
1. Rotary Pursuit; A vs. B	5.26[a]	1/726
2. Rudder Control; A vs. B	1.73	1/726
3. Complex Coordination; A vs. B	8.60[b]	1/726
4. Direction Control; A vs. B	0.47	1/274
Class Differences		
5. Rotary Pursuit; 1 through 16	0.67	15/726
6. Rudder Control; 1 through 16	1.11	15/726
7. Complex Coordination; 1 through 16	37.70[b]	15/726
8. Direction Control; 1 through 7	2.65[a]	6/274
Apparatus-by-Class Interaction		
9. Rotary Pursuit	0.66	15/726
10. Rudder Control	2.18[b]	15/726
11. Complex Coordination	1.51	15/726
12. Direction Control	1.19	6/274

[a]Difference reliable beyond .05 point.

[b]Difference reliable beyond .01 point.

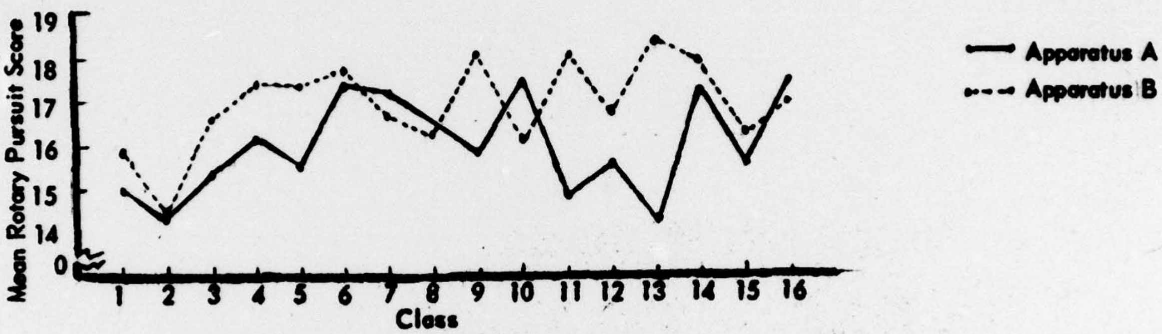


Fig. 1 Trends in mean Rotary Pursuit apparatus scores.

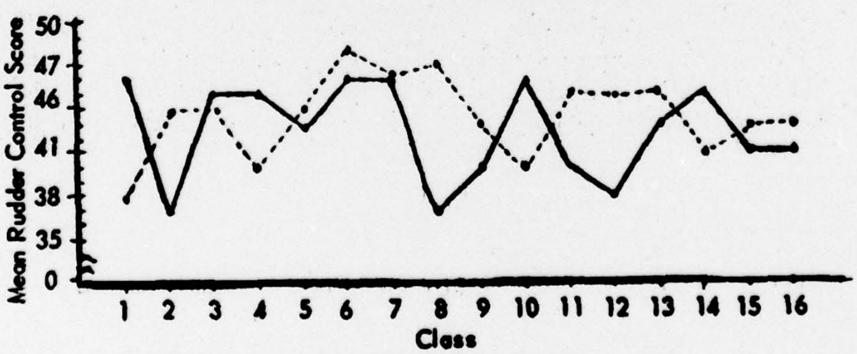


Fig. 2 Trends in mean Rudder Control apparatus scores.

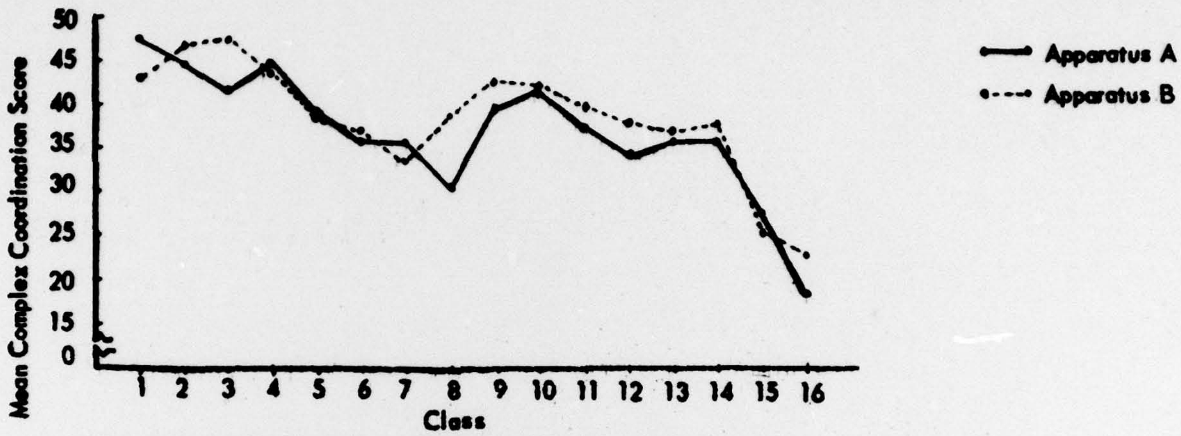


Fig. 3 Trends in mean Complex Coordination apparatus scores.

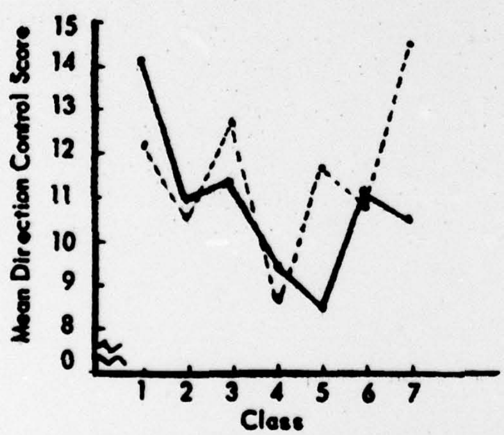


Fig. 4 Trends in mean Direction Control apparatus scores.

ROTARY PURSUIT

A significant difference was found between the means of Rotary Pursuit Apparatus A and Apparatus B, with the mean of Apparatus B significantly higher ($P < .05$). No significant variation in means was associated with class or apparatus-class interaction. The mean for Apparatus A was 15.90 and for Apparatus B, only 16.96--a small absolute difference (Table A-2). However, the standard deviations around the two means were sufficiently restricted to repudiate the hypothesis of random sampling from a common population.

RUDDER CONTROL

No significant differences between apparatus means or class means (Table A-3) were found for the Rudder Control Test. However, a significant apparatus-class interaction was found. The reason for this interaction can be seen most clearly in Figure 2. There was a decided tendency for class means to "cross-over"; that is, when half the class scored high on Apparatus A, the remainder of the class, tested on Apparatus B, tended to score low, and vice-versa.

COMPLEX COORDINATION

Highly significant differences between apparatus and class means were found for the Complex Coordination Test (Table A-4). Apparatus differences were small in absolute score terms. Means for scores on Apparatus A and B were 36.11 and 37.82, respectively. Class differences were much more dramatic (Figure 3), the mean scores dropping from 45.2 for class 1 to 20.5 for class 16. These effects were consistent over both copies of the apparatus and for all classes. Thus, no significant interaction was present.

DIRECTION CONTROL

No significant differences were found associated with apparatus or apparatus-class interaction for the Direction Control Test (Table A-5). A difference, significant at the .05 level, was found for mean scores associated with class. This difference was not caused by the more or less progressive decrement seen in the Complex Coordination Test, but a peculiar curvilinear effect whereby classes tested in the middle of the interval scored significantly lower than classes at the beginning and end of the interval.

IMPLICATIONS OF THE RESULTS

APPARATUS DIFFERENCES

Two of the four psychomotor tests studied, Rotary Pursuit and Complex Coordination, were found to generate significantly different mean scores between two copies of the same apparatus. Statistical significance is attributable to the relatively small within-group score dispersion and a large number of degrees of freedom for evaluation.

These results are similar to those found by the U. S. Air Force (Melton, 1947). The Air Force studies clearly showed that apparatus variance was a function of the particular apparatus test under consideration. The more complex the electro-mechanical design of the apparatus, the more marked were apparatus differences. The Air Force used two approaches to correct apparatus variance. One approach involved the development of rigorous and uniform calibration procedures and standards. The other approach involved assembling a single standard score conversion table for all testing stations.

The Air Force also conducted a series of studies comparing the validity of uncorrected raw scores and standard scores for prediction of graduation vs. elimination from pilot training. The results, based on thousands of cases, indicated that very little attenuation of the biserial validity coefficients occurred as a consequence of the use of the raw scores.

Where significant apparatus differences were found in the present study they indicated little practical significance. Considering this fact, and influenced by the Air Force validity studies, HFRB did not consider necessary a correction for apparatus differences in the Rotary Pursuit and Complex Coordination Tests.

CLASS DIFFERENCES

Two psychomotor tests, Complex Coordination and Direction Control, were found to have significantly different means associated with class, that is, with passage of time. The relatively progressive decrement shown in the mean scores of the Complex Coordination Test represents a serious psychometric problem. If cutting scores are established during a time interval when mean scores are high (or low), and future selection is made when mean scores are becoming progressively lower (or higher), the impact on input quotas and personnel quality may be serious. With serious decrement, a disproportionate number of personnel will be rejected; with serious increment, personnel quality may drop.

The Air Force also encountered usage variance in their apparatus tests. Even a relatively simple test, Finger Dexterity, showed usage variance due to the rounding off of pegs and holes. The more sophisticated Complex Coordination test provided particular problems because of the wearing of the contact plates which light the response lamps. This problem was also met by more effective calibration and maintenance standards.

Maintenance reports were not available for the machines used in the present study. To obscure the interpretation even further, the machines were moved from Fort Rucker, Alabama, to Camp Wolters, Texas, after the first nine classes was tested. Recalibration may have been accomplished before the ninth class was tested, a circumstance which may account for the temporary increase in mean scores shown for classes 9, 10, and 11. To test the hypothesis that the move itself or some factor associated with the move--humidity or maintenance, for example--was related to score variations, analyses of variance were computed which compared the means of the first nine classes tested at Fort Rucker with the remaining seven tested at Camp Wolters. Significant installation differences were found only for the Complex Coordination Test. However, Figure 3 shows that progressive decrease in mean scores occurred within the nine-class Rucker interval and within the seven-class

Wolters interval. It was concluded that class and apparatus differences were independent of the particular installation where testing was done.

Reasons for the curvilinearity of the Direction Control scores over classes are similarly unknown. Even though the total number of cases used in the analysis was less than the number used in the other three tests, the number was sufficiently high to yield stable variance estimates. With little first-hand knowledge of the testing or maintenance conditions, these results could not be interpreted.

In the Helicopter research program of the Human Factors Research Branch, the usage problem was met statistically by correcting Complex Coordination raw scores to standard T-scores for each class. No correction was considered necessary for the Direction Control Test.

CONCLUSIONS

Results of the study reemphasized the need for caution in the use of apparatus tests. The study reaffirmed the fact that individual differences in machines do occur and that scores tend to fluctuate significantly with continued use of the same machine. A statistical solution may be the only feasible means of completely neutralizing apparatus differences. Ideally, however, score fluctuations through continuing machine use should be handled by proper maintenance and not by statistical conversion.

REFERENCES

1. Fleishman, E. A. Psychomotor Selection Tests: Research and application in the United States Air Force. Personnel Psychol. 449-467, 1956.

2. Melton, A. W. (Ed.) Apparatus tests. AAF Aviation Psychology Program Research Reports, Vol. 4, 1947.

3. Zeidner, J., Martinek, Harold, and Anderson, Alan A. Evaluation of experimental predictors for selecting Army helicopter pilot trainees--II. HFRB Technical Research Note 101. December 1958.

APPENDIX

SUPPLEMENTARY TABLES OF DATA ON THE 16 U.S. ARMY PRIMARY HELICOPTER SCHOOL CLASSES IN THE STUDY

APPENDIX

Table A-1

PROPORTION PASSING THE HELICOPTER TRAINING PROGRAM

Class	Apparatus A N	Apparatus A P	Apparatus B N	Apparatus B P
1	20	.65	20	.35
2	19	.47	19	.42
3	24	.42	24	.58
4	25	.40	25	.56
5	18	.56	18	.44
6	27	.41	27	.52
7	11	.45	11	.55
8	13	.46	13	.77
9	12	.50	12	.67
10	28	.64	28	.54
11	33	.48	33	.39
12	36	.44	36	.39
13	35	.54	35	.54
14	20	.50	20	.40
15	27	.33	27	.41
16	31	.58	31	.68
TOTAL	379	.49	379	.50

APPENDIX

Table A-2

MEANS AND STANDARD DEVIATIONS FOR THE ROTARY PURSUIT TEST

Class	Apparatus A Mean	Apparatus A S.D.	Apparatus B Mean	Apparatus B S.D.
1	15.00	6.55	15.85	3.47
2	14.37	6.71	14.47	5.09
3	15.29	6.45	16.62	4.30
4	16.16	5.00	17.40	7.32
5	15.50	5.75	17.28	6.36
6	17.41	6.76	17.70	6.48
7	17.09	4.23	16.64	3.14
8	16.54	5.87	16.23	3.42
9	15.75	6.23	18.00	5.05
10	17.43	6.39	16.04	6.09
11	14.85	6.70	17.97	6.36
12	15.47	4.78	16.72	7.25
13	14.31	5.77	18.34	8.02
14	17.35	6.37	17.75	5.18
15	15.48	6.15	16.11	6.99
16	17.39	6.45	16.97	8.18
TOTAL	15.90	6.18	16.96	6.44

APPENDIX

Table A-3

MEANS AND STANDARD DEVIATIONS OF THE RUDDER CONTROL TEST

Class	Apparatus A		Apparatus B	
	Mean	S.D.	Mean	S.D.
1	46.25	11.94	37.55	10.71
2	37.00	10.43	44.05	11.76
3	45.25	11.29	44.00	10.23
4	45.76	8.59	40.52	13.97
5	42.61	8.45	43.89	13.24
6	45.93	10.29	48.19	9.17
7	45.73	11.35	46.45	12.36
8	37.15	9.90	47.15	9.31
9	39.67	11.89	43.17	13.53
10	45.61	10.91	40.04	11.60
11	39.70	12.31	45.03	9.85
12	38.19	11.62	44.64	10.38
13	43.37	9.05	45.37	9.52
14	44.75	9.74	40.85	7.03
15	40.96	12.71	42.70	10.01
16	40.87	12.60	42.65	13.59
TOTAL	42.42	11.41	43.49	11.35

APPENDIX

Table A-4

MEANS AND STANDARD DEVIATIONS FOR THE COMPLEX COORDINATION TEST

Class	Apparatus A Mean	Apparatus A S.D.	Apparatus B Mean	Apparatus B S.D.
1	47.40	8.48	43.05	7.45
2	44.74	7.59	46.89	9.76
3	41.54	11.41	47.71	7.52
4	44.44	9.88	43.72	7.29
5	38.28	5.38	38.06	6.50
6	35.89	6.26	36.78	6.39
7	35.73	5.71	33.64	7.80
8	30.23	8.37	38.92	5.81
9	39.58	8.60	42.42	9.10
10	41.54	6.62	41.82	8.41
11	37.58	8.92	39.76	8.90
12	34.08	9.45	37.64	7.65
13	35.43	6.38	36.74	8.27
14	35.40	9.54	37.55	5.90
15	26.70	6.61	25.81	6.57
16	18.16	5.48	22.87	7.75
TOTAL	36.11	10.87	37.82	10.22

APPENDIX

Table A-5

MEANS AND STANDARD DEVIATIONS OF THE DIRECTION CONTROL TEST

Class	Apparatus A Mean	Apparatus A S.D.	Apparatus B Mean	Apparatus B S.D.
1	14.10	6.17	12.15	5.50
2	10.89	4.99	10.58	5.83
3	11.38	5.85	12.75	6.19
4	9.44	6.16	8.60	4.90
5	8.50	5.56	11.67	5.93
6	11.00	5.30	10.93	5.05
7	10.58	3.06	11.32	5.69
TOTAL	10.86	5.77	11.32	5.79

APPENDIX

Table A-6

ANALYSIS OF VARIANCE OF ROTARY PURSUIT TEST DATA

Source	d.f.	M.S.	F	P
Apparatus	1	213.20	5.26	$< .05$
Class	15	27.17	----	N.S.
A X C	15	26.92	----	N.S.
Error	726	40.50		
TOTAL	757			

APPENDIX

Table A-7

ANALYSIS OF VARIANCE OF RUDDER CONTROL TEST DATA

Source	d.f.	M.S.	F	P
Apparatus	1	220.68	1.75	N.S.
Class	15	139.96	1.11	N.S.
A X C	15	276.02	2.18	$<.01$
Error	726	126.42		
TOTAL	757			

APPENDIX

Table A-8

ANALYSIS OF VARIANCE OF COMPLEX COORDINATION TEST DATA

Source	d.f.	M.S.	F	P
Apparatus	1	552.24	8.60	<.01
Class	15	422.14	37.70	<.01
A X C	15	96.92	1.51	N.S.
Error	726	64.24		
TOTAL	757			

APPENDIX

Table A-9

ANALYSIS OF VARIANCE OF DIRECTION CONTROL TEST DATA

Source	d.f.	M.S.	F	P
Apparatus	1	15.12	----	N.S.
Class	6	85.78	2.65	<.05
A X C	6	38.65	1.19	N.S.
Error	274	32.39		
TOTAL	287			